Data Quality Checklist

Data Collection

The purpose of data collection is to provide a basis for analysis, in other words, to turn data into information that can be used by, and useful to, decision-makers.

Data Quality

Data quality can be defined as "the extent to which information remains reliable and consistent across the organization". The issue of data quality often comes down to answering the question, "Is the data collected of use for the intended purpose?" Any data that is incorrect can potentially impact the quality of the data and any decision made from that data. Thus, poor quality data can have a negative impact on an organization since many management decisions are made based on quantitative analysis. Incomplete, inaccurate or missing data increases the risk of inaccurate reporting of findings and trend analysis. Moreover, having to scrub data to fix a particular problem is expensive and time consuming. Therefore, it pays to initiate a data quality system because improvements in data quality will lead to more informed management, strategic planning and decision-making.

Data Verification & Validation

To ensure that data collection and analysis is applicable to the decisions that will be made from the analysis, it is important to verify and validate the data. Verification ensures that the data collected represents the data desired – that a set of data meets a specified set of criteria. Validation is the comparison of the data with the requirements' specifications, i.e., does the data collected accurately represent the condition it is supposed to describe? Together, verification and validation check for any oversights or deviations from requirements, and identifies them.

ESS Annual Data Collection

All data within the ESS Annual Data online application are required under IDEA (Individuals with Disabilities Education Act). The application is broken out into multiple sections, with some data to be reported at the district level and some data to be reported at the school level. The areas of data collection include: Personnel, Exit, Discipline, IEP Evaluation Summary, and Performance Indicators.

Data Quality Checklist[1]

Regardless of the type of data collection system, whether electronic, manual or some other format, the following steps will help to ensure accuracy and reliability of data being reported:

1. **Validity.** Measure exactly what the objective describes. *For example, the number of personnel serving students with disabilities should be reported as a Full-Time Equivalency (FTE) in decimal format – not a head count and the FTE should be reported on or about December 1st of the reporting year – not during the entire school year. The age of students reported as exiting special education should be as of December 1 of the reporting year – not the age at the time of exit.* Is there a realistic plan in place to improve data validity and collection? Data can always be improved, and often, by making small system changes, the quality of data can improve significantly.

2. **Accurate Definitions.** Often terms that seem self-explanatory are not. *A simple example is the phrase "not fully certified". This data is collected within the Personnel section but do all LEAs define "not fully certified" the same way? Do LEAs include long-term substitutes? If each LEA uses their own definitions, this is of major concern. Why? Because the data reported may not accurately reflect a shortage of the actual FTE of personnel that are not fully certified, thus inaccurately representing the shortage of personnel that we have in our state.* No matter how detailed and comprehensive the definitions and instructions, not every definitional variation can be anticipated. *Additional space for LEAs to write in any deviations will be provided on the comments form, prior to the signature page*.

3. **Accurate Counts.** *If an LEA has the lowest dropout rate for five years and then that LEA jumps to the highest dropout rate, you would take a second look at the data, right?* In general, change takes time. That's why we recommend that you double-check the accuracy of your data if it jumps by more than 10% in either direction. Built-in over and under count checks are simple, effective and efficient ways to ensure the accuracy of your data. If you are collecting data by computer, simply build in validity checks. *For example, ensure that the totals discipline data reported by disability equal the total discipline data reported by ethnicity.* If you are using paper-reporting systems, simply outline a similar checking procedure. Ideally, you also want an independent check of the data; have someone outside the data collection process go through the same procedure. If your teachers did not complete your data request or survey *(i.e. the IEP Evaluation Summary information)*, it may extremely useful for you to find out why. Often it is because the survey was confusing, complicated, or had some other problem that made it prohibitive to complete. If you find out what the problems are, you can fix them.

[1] Currently in DRAFT form and designed by the US Department of Education, Planning & Evaluation Services Division

4. **Editing.** There are two easy ways to check your data:

   - Print out the data file. Scan for any uneven columns. Randomly check to make sure that the numbers are all in the appropriate range. *The Print Reports function within the online application is designed to assist you in this data check.*
   - Do what are called frequency distributions. These distributions show all the answers clustered together and make it easy to tell whether there are any numbers out of possible range. If a contractor is handling your data, ask to see and verify the data set using frequency distributions.

   Keep an eye out for unusual findings. Sometimes findings really are unusual, but always double check your data set first. When you do find a data error, and everyone does, make sure you find where they came from. By identifying these problems, you can address them and possibly change your system so they do not happen again. Just like you can build in checks to your electronic data collection, you can also build in checks to your overall database. For example, when disaggregating data, always check to make sure that those numbers, when aggregated, match the aggregated numbers in your database. To the extent possible, try to build in data editing checks. If you program checks into the software, it will mean that no one will forget to make that second check.

5. **Calculation.** Statistical software programs have different default ways to deal with missing data. Sometimes they automatically exclude missing values from your analysis, but sometimes they require you to specify that the missing values should be removed. In fact, you may find that missing values are actually included as actual values. This often happens when recoding or creating new variables. It is important to be mindful that missing data may cause problems. Be aware of how all the variables are coded and recoded. Ask for frequency distributions of all new variables and ask how missing values were handled.

6. **Timeliness.** Data are meaningful to the extent that they are relevant to the time period for which decisions are being made. Consider when your data collection should optimally occur and then schedule it appropriately. *ESS Annual Data is due annually every summer to ADE for the school year that recently ended. Alert your teachers at the beginning of the year that they need to maintain IEP data all year long. Have human resources flag personnel that serve students on or about December 1st. Track discipline data throughout the school year.* The focus is always on moving up the data quality continuum: how can we collect and use data more efficiently and effectively? The best time to think of improvements is while you are going through the process. Have your data collectors keep running notes on how to improve the systems. Give stakeholders rewards for suggesting improvements. Computer-accessible data comes up here again because electronic data collection usually means faster turn-around time. Data collection is a process that takes time and often cannot be accomplished after-the-fact. Even if you give out your data collection forms at the beginning of the school year, do not expect that people are reviewing the forms periodically to make sure they are tracking all the information that you need. Keep your

data collectors informed and reminded of all data requirements and their importance as well as collection timelines. Reminders throughout the funding cycle will only serve to refocus everyone on the importance of high-quality data.

7. **Full Disclosure.** If you were describing your data to a researcher in the field, how would you describe them? Whatever the potential issues with your data, you will want to report them wherever your data appears. Also, for the sake of improving your data, think about how those limitations could be addressed in revised future data collections. We collect this data so we can USE it. Did you get the data out to all the potential users? For example, the people who collected the data would benefit from a summary comparing them to others who collected similar data. Since data quality is often a chain (from the schools, to districts, to states, to OSEP), it is useful to make sure that there are opportunities for improvement to come from any part of that chain. You care about data quality or you would not have read this far. So, you probably have developed or will develop a data system that accounts for data quality issues. Make sure to document that system and have those documents accessible to everyone on your team.

Helpful Hints

In preparation for collecting the required data for the ESS Annual Data Collection, be sure to read the instructions for all the reporting sections, take the time to review the definitions in the data dictionary and hyperlinks found throughout the application, check out the Frequently Asked Questions (FAQ), attend workshops and seek clarification from ADE/ESS Data Management.